CS235 Project Proposal

Project Type: Software Application
Team Members: Gowtham Tumati
Sudip Bala
Lovepreet Singh Dhaliwal
Abhishek Ayachit
Labor Division:

1. Time Series Analysis for Prediction
2. To check if the stock makes profit for current data based on historical data
3. To check if the stock makes profit for the future data that is predicted.

The team would be divided into groups of two which would perform two different algorithms for the above three use cases.
Groups:

1. Abhishek Ayachit & Sudip Bala
2. Lovepreet Singh Dhaliwal & Gowtham Tumati

Evaluation Plan: A range of data will be kept for training purposes and remaining data is used as a test dataset.
Dataset: S&P 500 Stock Data. (https://www.kaggle.com/camnugent/sandp500#ADBE_data.csv)
Evaluation Criteria: The algorithm that is trained will be tested against dataset which would be a small range of data from the above mentioned dataset.

Description
Motivation:
Stock market analysis is an attempt to understand the current and historical trends in order to determine the future of a particular stock or the entire stock market. Stock analysis helps different stock market investors and traders to make important decision regarding buying and selling of the stocks. Most of the traders rely only on the recent data, their intuition and the experience gained over the years. But they don't use the factual data available to them to make these decision. The data about the stock market is available but the data needs to be computationally analyzed and then the meaningful, data supported decisions can be made. Therefore, this will give an advantage to the investors and traders over the others.

Method:
The dataset we are using, i.e., S&P 500 Stock Data, consists of following fields:
Date : Open : High : Low : Close : Volume : Name

- A web scrapper will be used in order to gather the data about the stock market from the web. Then the following mentioned algorithms can be used to make the prediction the value and to check if the stock will make profit or not. To simplify the process, for now the available dataset which is mentioned above (S&P 500 Stock Data) will be used. The web scrapper will be used, if time permits, in the final part of the project in order analyze the data in real time.

- Autoregression Model:
  This module takes the data to analyze it and based on this, provides the prediction for future cases for which the prices of the stocks need to be decided. . For the analysis of the stock

market to predict the stock price, the method we are going to use is Autoregressive Method(AR).

Autoregression Method(AR):

- The autoregression(AR) method models the next steps in the sequence as a linear function of the observations at prior steps.
- The reason we are choosing this model is that the method is suitable for univariate time series without trend and seasonal components.

Libraries required for analysis (Python):
NumPy: provides the functions to work with arrays and matrices.
Pandas: to easily access and manipulate numerical and time series data.
Statsmodels: to explore the data and perform statistical analysis and estimate statistical models and execute the statistical tests.

- Moving Average Model (MA):
  The moving average models the next step in the sequence as a linear function of the residual errors from a mean process at prior time steps.

  As this method is used for the same reason as mentioned in the case of Autoregression or AR model, the only difference is that this method uses the errors of the residuals.

Intended experiments:

- To check the correctness of the prediction, a certain range of data can be kept aside and when the algorithm is trained with the rest of the data, this data can be used by running the algorithm for this set of data values.
  - Following Metrics will be used to evaluate the prediction models:
    Mean Absolute Error (MAE): it is the mean of the absolute value of the errors
    Mean Squared Error (MSE): it is the mean of the squared errors
    Root Mean Squared Error (RMSE): it is the square root of the mean of the squared errors
  - For the real time application, if the data is not sufficient, Cross-validation can be used to check the accuracy of the algorithm

- To check the correctness of the second use case mentioned above, following method will be used. If the actual value is greater than the current value and the predicted value is also greater than the current value, then the stock is said to be in the profit. Similarly, for the loss, predicted value can be calculated. This is how the correctness of the profit and loss module will be checked.
  Even if the value of the predicted value is not even near to the actual value, it won't be a big factor as the algorithm parameters can be tweaked a little bit to get the correct value.

Future Work:
The project can be extended to collect the data everyday as the stock market closes and based on the previous data and that day's data, predict the stock values for the next day.
If time permits, this will be second part of the project.